Công thức xấp xỉ

- **Phân phối Siêu bội**
  Xét tập hợp gồm N phần tử trong đó có M phần tử có tính chất $\wp$.
  Chọn *không hoàn lại* ngẫu nhiên n phần tử và gọi X là số phần tử có tính chất $\wp$ trong n phần tử chọn ra, thì X có phân phối siêu bội, ký hiệu $X \sim H(N, M, n)$.
  $$P(X = x) = \frac{C_M^x C_{N-M}^{n-x}}{C_N^n}.$$
- **Phân Phối Nhị thức**
  Xét tổng thể (kích thước rất lớn) mà tỷ lệ phần tử có tính chất A là p. Chọn *không hoàn lại* ngẫu nhiên n phần tử. Gọi X là số phần tử có tính chất A trong n phần tử. Khi đó X có phân phối Nhị thức.
  $$X \sim B(n, p)$$
  $$P(X = k) = C_n^k p^k (1-p)^{n-k}, \qquad x = 0, 1, 2, \ldots, n$$

Xấp xỉ phân phối Siêu bội bằng phân phối Nhị thức

- Điều kiện: n rất nhỏ so với N
- Cụ thể: $n \leq 0.01N$
- Khi đó: $H(N, M, n) \longrightarrow B(n, p = M/N)$

VD. Một vườn lan có 10.000 cây sắp nở hoa, trong đó có 1.000 cây hoa màu đỏ.
1) Tính xác suất để khi chọn ngẫu nhiên 20 cây lan thì được 5 cây có hoa màu đỏ.
2) Tính xác suất để khi chọn ngẫu nhiên 50 cây lan thì được 10 cây có hoa màu đỏ.
3) Có thể tính xác suất để khi chọn ngẫu nhiên 200 cây lan thì có 50 cây hoa màu đỏ được không ?

Liên hệ giữa Phân phối Nhị thức và Phân phối Poisson

1) Một ĐLNN X được gọi là có **phân phối Poisson** với tham số λ, kí hiệu $X \sim P(\lambda)$, nếu các giá trị của nó là các số nguyên không âm, và có hàm khối xác suất

$$P(x) = P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad \text{với } x = 0, 1, 2, \ldots$$

2) Nếu ĐLNN X có phân phối Nhị thức, $X \sim B(n, p)$ thì

$$P(X = x) = C_n^x p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n$$

3) Với mỗi $x = 0, 1, 2, \ldots$, ta có giới hạn sau:

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\lambda}} C_n^x p^x (1-p)^{n-x} = e^{-\lambda}\frac{\lambda^x}{x!}$$

Nhị thức xấp xỉ về Poisson

$$\boxed{X \sim B(n, p) \;\Rightarrow\; X \sim P(\lambda = np)}$$

- Điều kiện: n khá lớn và p khá nhỏ.
- Cụ thể: $p \leq 5\%$, $n \geq 30$ và $np < 5$.

$$\boxed{P(X = x) \approx e^{-\lambda}\frac{\lambda^x}{x!}}$$

Ví dụ 1

Last month your company sold 1,000 new watches. Past experience indicates that the probability that a new watch will need repair during its warranty period is 0.002. Compute the probability that:

1. *no watch will need warranty work.*
2. *no more than 5 watches will need warranty work.*

Tính chính xác bằng phân phối nhị thức:

Gọi X là số đồng hồ bị hư trong thời gian bảo hành (trong tổng số 1000 đồng hồ công ty sản xuất), suy ra $X \sim B(n, p)$ với

- $n = 1000$ đồng hồ;
- $p =$ xác suất 1 đồng hồ bị hư trong thời gian bảo hành $= 0.002$

1. Xác suất không có đồng hồ nào bị hỏng:

$$P(X = 0) = f(0) = 13.51\%$$

2. Xác suất có *không quá* 5 đồng hồ bị hỏng:

$$P(X \leq 5) = F(5) = 98.35\%$$

Tính xấp xỉ bằng phân phối Poisson:

Vì $p \leq 5\%$ và $n > 30$, nên ta có thể xấp xỉ từ phân phối nhị thức về phân phối Poisson: $X \sim P(\lambda = np = 2)$.

1. Xác suất không có đồng hồ nào bị hỏng:

$$P(X = 0) = f(0) = 13.53\%$$

2. Xác suất có *không quá* 5 đồng hồ bị hỏng:

$$P(X \leq 5) = F(5) = 98.34\%$$

Theorem 1 (Central Limit Theorem)

Cho $X_1, X_2, \ldots, X_n$ là các BNN độc lập có cùng kỳ vọng μ và độ lệch chuẩn σ. Đặt

$$S_n = X_1 + X_2 + \ldots + X_n.$$

Đặt $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$. *Khi* $n \longrightarrow \infty$, *ta có*

$$P(Z_n \leq z) \longrightarrow \Phi(z) \quad \forall\ z$$

Ý nghĩa: Khi n rất lớn, ta có:

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < z\right) \sim \Phi(z)$$

với $\Phi(z)$ là hàm Gauss.

Xấp xỉ pp nhị thức về pp chuẩn

Hệ quả 1

Khi n đủ lớn và p không gần 0, không gần 1, phân phối Nhị thức $B(n,p)$ có thể xấp xỉ về phân phối Chuẩn $N(\mu,\sigma^2)$, với $\mu = np$ và $\sigma = \sqrt{np(1-p)}$.

Trong thực tế, người ta hay chọn điều kiện xấp xỉ:

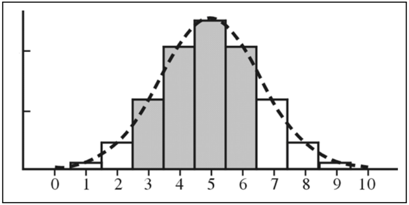
$$n > 30, \quad np \geq 5, \quad n(1-p) \geq 5, \quad 0.05 \leq p \leq 0.95$$

$$\boxed{B(n,p) \approx N(\mu,\sigma^2), \quad \mu = np, \ \sigma = \sqrt{np(1-p)}}$$

Điều chỉnh 0.5

Chú ý. khi xấp xỉ phân phối nhị thức *(giá trị nguyên rời rạc)* bởi phân phối chuẩn *(giá trị thực liên tục)* ta cần điều chỉnh 0.5

- $P(a \leq X \leq b) \approx P(a - 0.5 < X < b + 0.5)$
- $P(X \leq b) \approx P(X < b + 0.5)$
- $P(X < b) \approx P(X < b - 0.5)$
- $P(a \leq X) \approx P(a - 0.5 < X)$
- $P(X = a) \approx P(a - 0.5 < X < a + 0.5)$



Điều chỉnh 0.5

Ví dụ 2

Cho $X \sim B(100, 0.4)$*, tính các xác suất*

1. $P(20 \leq X \leq 40)$
2. $P(X < 32)$
3. $P(X \geq 25)$

Hãy xấp xỉ X về phân phối chuẩn và tính lại các xác suất trên.

Vì $np = 40, nq = 60 > 5$, nên ta có thể xấp xỉ $B(n = 100, p = 0.4)$ về phân phối chuẩn $N(\mu, \sigma^2)$, với $\mu = np = 60$ và $\sigma = \sqrt{npq} = 2\sqrt{6}$.

Example 4.15. A new computer virus attacks a folder consisting of 200 files. Each file gets damaged with probability 0.2 independently of other files. What is the probability that fewer than 50 files get damaged?

Solution. The number X of damaged files has Binomial distribution with $n = 200$, $p = 0.2$, $\mu = np = 40$, and $\sigma = \sqrt{np(1-p)} = 5.657$. Applying the Central Limit Theorem with the continuity correction,

$$\begin{aligned} \boldsymbol{P}\{X < 50\} &= \boldsymbol{P}\{X < 49.5\} = \boldsymbol{P}\left\{\frac{X-40}{5.657} < \frac{49.5-40}{5.657}\right\} \\ &= \Phi(1.68) = \underline{0.9535}. \end{aligned}$$